تحسين الصيانة التنبؤية لآلات التفريز باستخدام هندسة الميزات ونماذج تعلّم الآلة الهجينة

إعداد : صفاء مصطفى محسن إشراف: د. حسام الوفائي

الملخص:

يركّز هذا البحث على تطوير نموذج صيانة تنبؤية يعتمد على تقنيات تعلّم الآلة بهدف تحسين موثوقية وكفاءة آلات التفريز في البيئات الصناعية الحديثة، تعتمد المنهجية على تحليل بيانات تشغيلية معيارية من قاعدة AI4I2020، ومعالجة التحديات المرتبطة بالصيانة التقليدية مثل التوقفات المفاجئة وارتفاع تكاليف التشغيل. يتضمن النموذج المقترح مراحل تنظيف البيانات واستخراج ميزات مشتقة تعبّر عن السلوك التشغيلي للآلة، ثم تدريب نماذج تعلم الآلة مثل Random Forest وXGBoost وGradient Boosting، إضافة إلى تصميم شبكة عصبونية مخصّصة للتعامل مع الأعطال النادرة.

تُظهر النتائج أن استخدام الميزات المشتقة حسّن أداء النماذج بشكل ملحوظ، خصوصاً في ، مقارنة بالاعتماد على الميزات الخام فقط، كما حققت F1-Scoreو Recall مؤشرات النماذج الشجرية التجميعية أداءً قوياً في تصنيف الأعطال الشائعة، بينما تطلّب اكتشاف استخدام تقنيات معالجة عدم التوازن مما رفع قدرة النموذج على TWF الأعطال النادرة مثل التعرّف على هذه الحالات رغم محدودية عيناتها، وتؤكد الدراسة أن دمج هندسة الميزات مع نماذج هجينة مخصّصة للفئات النادرة يوفّر إطاراً فعّالاً لبناء نظام صيانة تنبؤية قابل للتطبيق الصناعي.

كلمات مفتاحية: الصيانة التنبؤية، تعلّم الآلة، آلات التفريز، ميزات مشتقة، AI4I 2020.

Enhancing Predictive Maintenance of Milling Machines Using Feature

Abstract:

This research focuses on developing a predictive maintenance model based on machine learning techniques to enhance the reliability and operational efficiency of milling machines in modern industrial environments. The methodology relies on analyzing standardized operational data from the AI4I 2020 dataset and addressing challenges associated with traditional maintenance approaches, such as unexpected failures and high operational costs. The proposed model includes data cleaning, the extraction of engineered features that capture the machine's operational behavior, and the training of machine learning algorithms such as Random Forest, XGBoost, and Gradient Boosting, in addition to designing a neural network specifically tailored to handle rare failure cases.

The results show that incorporating engineered features significantly improved model performance, particularly in Recall and F1-Score metrics, compared to relying solely on raw features. Ensemble tree-based models demonstrated strong performance in classifying common failure modes, while detecting rare failures such as TWF required specialized imbalance-handling techniques, including RandomOverSampler, Focal Loss, and pos_weight, which enhanced the model's ability to identify these infrequent cases despite their limited representation.

The study confirms that combining feature engineering with hybrid models tailored for rare classes provides an effective framework for building an industrially applicable predictive maintenance system. It further recommends developing continuous monitoring mechanisms for the model, improving data quality, and expanding datasets related to rare failures to strengthen system reliability and reduce unplanned downtime**.**

Keywords: Predictive Maintenance, Machine Learning, Milling Machines, Engineered Features, AI4I 2020.

مقدمة:

تُعدّ الصناعات التحويلية من أهم القطاعات الداعمة للنمو الاقتصادي العالمي، نظراً لاعتمادها على منظومات إنتاجية متقدمة قادرة على تحويل المواد الخام إلى منتجات عالية القيمة، وقد أكدت الدراسات الحديثة أن التطور في تقنيات التصنيع الدقيقة أصبح عنصراً محورياً في تعزيز القدرة التنافسية للدول الصناعية، خاصة مع تزايد الطلب على منتجات ذات جودة عالية ومعايير هندسية صارمة [1].

وتُعد عمليات التشغيل الميكانيكي، وعلى رأسها التفريز، من الركائز الأساسية في تصنيع المكوّنات الدقيقة المستخدمة في قطاعات الطيران والسيارات والروبوتات والأجهزة الطبية.

تحتل آلات التفريز مكانة متقدمة بين آلات التشغيل، لما تتمتع به من مرونة عالية وقدرة على تشكيل المواد الصلبة وإنتاج أشكال هندسية معقدة، وتشير الدراسات الحديثة إلى أن التطور في أنظمة التحكم، وتحسين ديناميكية الماكينات، واعتماد تقنيات المراقبة الذكية، جعل آلات التفريز أكثر موثوقية وكفاءة في بيئات التصنيع المتقدمة [1].

ومع تزايد تعقيد الأنظمة الإنتاجية، أصبحت الأعطال المفاجئة لآلات التشغيل من أبرز التحديات التي تواجه المصانع الحديثة، لما تسببه من خسائر مالية كبيرة وتوقفات غير مخططة، وقد بينت دراسات صناعية حديثة أن التوقفات غير المخططة تمثل ما بين 15% و25% من إجمالي خسائر الإنتاج، مما يجعل إدارة الأعطال وتحسين موثوقية المعدات ضرورة استراتيجية [2].

وفي ظل هذا الواقع، برزت الصيانة التنبؤية كأحد أهم الاتجاهات الحديثة في الصناعة، نظراً لقدرتها على الانتقال من مفهوم الصيانة التقليدية القائمة على الجداول الزمنية أو التدخل بعد حدوث العطل، إلى نموذج أكثر ذكاءً يعتمد على تحليل البيانات التشغيلية للآلة بهدف التنبؤ بالأعطال قبل وقوعها.

كما أثبتت الدراسات الحديثة أن دمج الصيانة التنبؤية مع تقنيات تعلّم الآلة وإنترنت الأشياء الصناعي يعزز دقة التنبؤ بالأعطال ويجعل أنظمة الإنتاج أكثر كفاءة واستجابة لمتطلبات الصناعة الحديثة.

أن الصيانة التنبؤية تسهم في تقليل التوقفات غير المخططة بنسبة قد تصل إلى 40%، وتحسين كفاءة التشغيل بشكل كبير [3].

أثبتت الدراسات الحديثة أن خوارزميات تعلم الآلة تُظهر أداءً عالياً في مهام تشخيص الأعطال، خاصة في بيئات التشغيل المعقدة مثل آلات التفريز، كما أن دمج تقنيات تعلّم الآلة مع بيانات الاستشعار اللحظية يساهم في بناء نماذج تنبؤية أكثر دقة وموثوقية، مما يعزز من فعالية أنظمة الصيانة التنبؤية في الصناعة الحديثة [4].

وانطلاقاً من هذه التطورات، يهدف هذا البحث إلى توظيف تقنيات تعلّم الآلة في تحليل بيانات التشغيل الخاصة بآلات التفريز، بهدف التنبؤ بالأعطال وتحسين موثوقية الأنظمة الإنتاجية، بما يتماشى مع توجهات الصناعة الحديثة نحو التحول الرقمي ورفع كفاءة التشغيل.

تنظيم البحث:

تم تنظيم البحث وفق الترتيب التالي: تم البدء بمقدمة عن العمل المنجز ثم تم ذكر الهدف من البحث وطرائقه، وبعدها تم إجراء دراسة مرجعية في الفقرة 1 تستعرض أحدث الأبحاث التي تم إنجازها عن الصيانة التنبؤية، ثم تم في الفقرة 2 شرحاً تفصيلياً لآلية بناء النموذج المقترح، بدءاً من معالجة البيانات وتنظيفها واستخراج الميزات المشتقة، وصولاً إلى تدريب نماذج تعلّم الآلة المختلفة مثل Random Forest وXGBoost وGradient Boosting، إضافة إلى تصميم شبكة عصبونية مخصّصة للتعامل مع الأعطال النادرة وأخيراً تم مناقشة النتائج التي تم الحصول عليها في الفقرة 3 وتحليلها مع مقارنة هذه النتائج مع الأبحاث السابقة في الفقرة 4.

هدف البحث:

يهدف هذا البحث إلى تطوير نموذج صيانة تنبؤية يعتمد على تقنيات تعلّم الآلة والشبكات العصبونية للتنبؤ بالحالات التشغيلية والأعطال المحتملة في آلات التفريز، وذلك من خلال تحليل بيانات قاعدة AI4I 2020 وتقييم أداء النماذج المقترحة وملاءمتها لمتطلبات البيئات الصناعية الحديثة.

مواد وطرق البحث:

يعتمد هذا البحث على تحليل بيانات تشغيلية معيارية من قاعدة AI4I 2020 باستخدام منهجيات تعلم الآلة، تمت معالجة البيانات واستخراج الميزات المشتقة وتدريب النماذج تمّ عبر Jupyter Notebook مع بيئة البرمجة Python ومكتباتها المتقدمة مثل Pandas و Scikit-Learnو PyTorch.

1 دراسة مرجعية

تناولت العديد من الدراسات الحديثة موضوع الصيانة التنبؤية باستخدام تقنيات تعلّم الآلة، وقد اختلفت هذه الدراسات في نوع البيانات المستخدمة، الخوارزميات، ومستوى النضج التطبيقي.

حيث قام (Akyaz, T., & Engin) في [5] بتقديم دراسة لنظام صيانة تنبؤية لآلات الغزل الصناعية باستخدام خوارزميات مثل Random Forest وXGBoost وSVM والشبكات العصبونية، وتمثلت قوة الدراسة في وجود إطار تطبيقي متكامل يربط بين جمع البيانات وتحليلها وتنفيذ النموذج في بيئة صناعية فعلية، إضافة إلى استخدام مجموعة واسعة من الميزات التشغيلية، إلا أن الدراسة لم تقدّم معالجة متقدمة لمشكلة عدم توازن البيانات، كما لم تتضمن هندسة ميزات مشتقة أو تحليلاً تفسيرياً عميقاً للمتغيرات المؤثرة في الأعطال.

بينما استخدم (Baradaran, A. H, 2025) في [6] مجموعة من نماذج التعلم الخاضع للإشراف مثل Naive Bayes وSVM والانحدار اللوجستي وRandom Forest لتشخيص حالة المحركات الكهربائية، وتميزت الدراسة بشمولية المقارنة بين عدة خوارزميات تقليدية وسهولة تفسير نتائجها، إلا أنها اعتمدت على نماذج تقليدية فقط دون استخدام شبكات عصبونية أو نماذج هجينة، كما لم تعالج مشكلة عدم توازن البيانات، مما قلل من قدرتها على اكتشاف الأعطال النادرة، إضافة إلى غياب إطار تطبيقي متكامل.

أما (Waghulde et al, 2025) في [7] فقد ركز على تقييم أداء مجموعة واسعة من نماذج تعلم الآلة والتعلم العميق باستخدام قاعدة بيانات AI4I 2020 ، بما في ذلك Random Forest وSVM وXGBoost والشبكات العصبونية، وتمثلت قوة الدراسة في شمولية المقارنة وتغطية عدد كبير من الخوارزميات، إضافة إلى استخدام مقاييس أداء متعددة، ومع ذلك، اقتصرت الدراسة على التقييم فقط دون تقديم إطار تطبيقي متكامل، كما لم تتناول هندسة ميزات مشتقة أو تقنيات متقدمة لمعالجة الأعطال النادرة، مما يجعل نتائجها أقل قابلية للتطبيق الصناعي المباشر.

بينما (Zhao et al, 2025) في [1] قام بتطوير نموذج صيانة تنبؤية لآلة التفريز باستخدام خوارزميات تعلم الآلة مدعومة بتحليل SHAP لتفسير قرارات النموذج، وتمثلت قوة الدراسة في قدرتها على تحديد المتغيرات الأكثر تأثيراً في حدوث الأعطال، مما يعزز قابلية تفسير النموذج، إلا أن الدراسة اعتمدت على الميزات الخام فقط دون هندسة ميزات مشتقة، كما لم تقدّم معالجة كافية لمشكلة عدم توازن البيانات، مما يحدّ من قدرتها على اكتشاف الأعطال النادرة بدقة.

وتُظهر هذه الدراسات مجتمعة أهمية توظيف تقنيات تعلم الآلة في بناء أنظمة صيانة تنبؤية فعّالة، إلا أنها تختلف في منهجياتها من حيث نوع البيانات، الخوارزميات المستخدمة، وعمق معالجة مشكلة عدم التوازن.

كما أن معظمها يعتمد على الميزات الخام دون هندسة ميزات مشتقة أو نماذج مخصصة للأعطال النادرة، وهي فجوات يسعى هذا البحث إلى معالجتها من خلال تطوير نموذج أكثر

تكاملاً يعتمد على هندسة الميزات وتقنيات معالجة عدم التوازن ضمن إطار تطبيقي مخصص لآلات التفريز.

2 بناء النموذج التنبؤي المقترح:

2.1 البيانات المستخدمة كأساس لبناء النموذج المقترح:

اعتمدت هذه الدراسة على قاعدة البيانات المرجعية AI4I 2020 بوصفها المصدر الرئيسي لبناء وتدريب النموذج التنبؤي المقترح، نظراً لكونها من أكثر قواعد البيانات استخداماً في أبحاث الصيانة التنبؤية.

تُعد من قواعد البيانات الصناعية المصممة لتمثيل السلوك التشغيلي لآلات التفريز تحت ظروف تشغيل مختلفة، بما في ذلك الحالات الطبيعية وحالات التدهور والأعطال [8].

وتتألف القاعدة من نحو 10,000 عينة، وهو حجم مناسب لتدريب نماذج تعلّم الآلة دون التعرض لمشكلة فرط التعميم، تتضمن قاعدة البيانات:

- UDI : رقم تعريف فريد لكل سجل
- Product ID : يتضمن رمز جودة المنتج
- Type : نوع المنتج
- Air temperature [K] : درجة حرارة الهواء
- Process temperature [K] : درجة حرارة العملية
- Rotational speed [rpm]: سرعة الدوران
- Torque [Nm]: العزم
- Tool wear [min] : تآكل الأداة
- TWF (Tool Wear Failure) : عطل ناتج عن تآكل الأداة

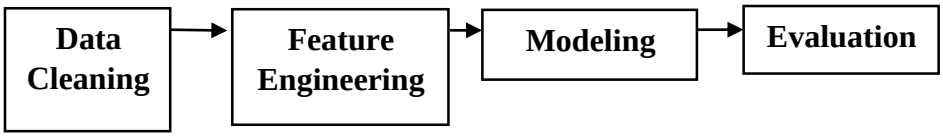
- HDF (Heat Dissipation Failure) : عطل في تبديد الحرارة
- PWF (Power Failure) : عطل في القدرة الكهربائية
- OSF (Overstrain Failure) : عطل بسبب الإجهاد الزائد
- RNF (Random Failure) : عطل عشوائي
- Machine failure: عطل الآلة

يُعد التوزيع غير المتوازن بين الفئات أحد أبرز التحديات في هذه القاعدة، حيث تشكل الحالات السليمة النسبة الأكبر من البيانات، بينما تظهر الأعطال النادرة وخاصة TWF بنسب منخفضة جداً مقارنة ببقية الفئات. يؤدي هذا التفاوت إلى انحياز النماذج نحو الفئات الشائعة، ويضعف قدرتها على اكتشاف الأعطال النادرة [9].

وبناءً على ما سبق، تشكل قاعدة AI4I 2020 أساساً قوياً لتطوير نموذج صيانة تنبؤية يعتمد على هندسة الميزات ومعالجة عدم التوازن، وهو ما سيتم تفصيله في القسم التالي ضمن منهجية بناء النموذج المقترح.

2.2 المنهجية العملية لتصميم نموذج الصيانة التنبؤية:

يوضّح شكل 1 مخطط سير العمل المعتمد في بناء النموذج التنبؤي المستخدم في هذه الدراسة، والذي يتكوّن من أربع مراحل رئيسية: تنظيف البيانات، واستخراج الميزات، وبناء النماذج، ثم التقييم.



شكل 1 المنهجية العملية لتصميم نموذج الصيانة التنبؤية

- معالجة البيانات وتنظيفها

تُعدّ معالجة البيانات خطوة أساسية في بناء أي نموذج يعتمد على تقنيات تعلّم الآلة، إذ تؤثر جودة البيانات وتنظيمها بشكل مباشر في دقة النموذج وقدرته على التعميم، وبناءً على ذلك، تم اتباع منهجية منهجية دقيقة لمعالجة البيانات التشغيلية المستمدة من قاعدة AI4I 2020 قبل البدء بمرحلة التدريب.

بدأت عملية المعالجة بإجراء تحليل أولي للبيانات بهدف فهم طبيعتها وتوزيعها واكتشاف أي قيم شاذة أو غير منطقية قد تؤثر في أداء النماذج.

بعد ذلك، تم تنفيذ ترميز للمتغيرات الفئوية مثل نوع المنتج باستخدام أسلوب الترميز العددي، وذلك لضمان إمكانية استخدامها ضمن نماذج تعلّم الآلة التي تتطلب مدخلات رقمية، كما تم تطبيق التطبيع بهدف تقليل الفروقات الكبيرة في نطاق القيم وتحسين استقرار عملية التدريب، خاصة في النماذج الحساسة لمقاييس البيانات مثل الشبكات العصبونية، وفي خطوة لاحقة، تم تقسيم البيانات إلى مجموعتي تدريب واختبار بنسبة مناسبة لضمان تقييم موضوعي لأداء النماذج [10].

أسهمت هذه الخطوات في تجهيز البيانات بصورة مثالية لبناء النموذج التنبؤي، وضمان أن تكون المدخلات نظيفة، متسقة، ومهيأة لعمليات هندسة الميزات ومعالجة عدم التوازن التي تشكل المرحلة التالية في منهجية بناء النموذج المقترح.

- هندسة الميزات

تُعدّ هندسة الميزات خطوة محورية في بناء النموذج التنبؤي، إذ تعتمد دقة النماذج وفعاليتها على جودة التمثيل الرقمي للسلوك التشغيلي للآلة، فالاعتماد على الميزات الخام وحدها غالباً لا يكفي لالتقاط الأنماط الدقيقة المرتبطة بالأعطال، خصوصاً الأعطال النادرة التي تتطلب توصيفاً

أعمق للسلوك الديناميكي للآلة، ومن هذا المنطلق، تم في هذه الدراسة تطوير مجموعة من الميزات المشتقة التي تعزز قدرة النماذج على التمييز بين الحالات السليمة وحالات العطل.

تمثل هذه الميزات المشتقة أحد العناصر الأساسية في أنظمة الصيانة التنبؤية، إذ تُستخلص من المتغيرات التشغيلية اعتماداً على العلاقات الفيزيائية والوظيفية المرتبطة بأنماط الأعطال المختلفة، وتتيح هذه الميزات توصيفاً كمياً أكثر دقة للحالة الصحية لآلة التفريز تحت ظروف تشغيل وتحميل متباينة، مما يسهم في تحسين قدرة النماذج على اكتشاف التدهور المبكر والتنبؤ بالأعطال قبل وقوعها [11].

- تدريب نماذج تعلّم الآلة

بعد تجهيز البيانات ومعالجتها واستخراج الميزات المشتقة ومعالجة مشكلة عدم التوازن، تم الانتقال إلى مرحلة تدريب نماذج تعلّم الآلة بهدف بناء نظام قادر على التنبؤ بالأعطال التشغيلية في آلات التفريز بدقة وموثوقية.

وقد تم اختيار مجموعة من الخوارزميات الشجرية التجميعية وخوارزميات التعزيز التدريجي نظراً لقدرتها العالية على التعامل مع البيانات الصناعية المعقدة، إضافة إلى أدائها المتميز في مهام التصنيف متعددة الفئات [12].

- تصميم الشبكة العصبونية المخصّصة للأعطال النادرة

نظراً للتحديات التي تواجه نماذج تعلّم الآلة التقليدية في اكتشاف الأعطال النادرة، تم في هذه الدراسة تصميم شبكة عصبونية اصطناعية مخصّصة لمعالجة هذه الفئة من الأعطال [13].

تتكون الشبكة لدينا من:

❖ طبقة الإدخال(Input Layer)

هي الطبقة التي تســتقبل جميع الميزات بعد تجهيز البيانات، ويكون عدد خلاياها مســاوياً لعدد الميزات.

❖ الطبقة المخفية الأولى (Hidden Layer 1 — 64 Neurons)

هي طبقة كثيفة تحتوي على 64 خلية عصبونية، وتعمل على تعلّم العلاقات الواسعة بين الميزات والتقاط الأنماط غير الخطية وبناء تمثيل أولي للبيانات.

❖ الطبقة المخفية الثانية (Hidden Layer 2 — 32 Neurons)

تتكون من 32 خلية عصبونية، وتهدف إلى ضغط المعلومات وتقليل الضجيج واستخراج الأنماط الأكثر أهمية [14].

❖ الطبقة المخفية الثالثة (Hidden Layer 3 — 16 Neurons)

تضم هذه الطبقة 16 خلية عصبونية، وتعمل على تعلّم الأنماط الدقيقة المرتبطة بالعطل والتقاط العلاقات الصــــغيرة التي قد لا تظهر في الطبقات الســـابقة، كما تهيئ البيانات لطبقة الإخراج باعتبارها مرحلة فلترة نهائية قبل اتخاذ القرار.

وتعتمد جميع الطبقات المخفية على دالة التنشـــيط ReLU، نظراً لفعاليتها العالية في تمثيل العلاقات غير الخطية وسرعة حسابها [15].

❖ طبقة الإخراج (Output Layer)

يختلف تصميمها بحسب نوع المهمة، ففي حالة التصنيف الثنائي (1/0) تُستخدم خلية واحدة مع دالة التنشيط Sigmoid التي تحوّل المخرجات إلى قيمة بين 0 و1، مما يجعلها مناسبة للنماذج الاحتمالية التي تهدف إلى تحديد احتمال حدوث العطل.

أما في حالة التصنيف متعدد الفئات (0–5) فتتكون طبقة الإخراج من ست خلايا، وتُستخدم دالة التنشيط Softmax التي تحوّل المخرجات إلى توزيع احتمالي يمثّل احتمال انتماء العينة إلى كل فئة، بحيث يكون مجموع الاحتمالات مساوياً لـ 1 [16].

وتتمثل الوظيفة الأساسية لطبقة الإخراج في إنتاج الاحتمال النهائي للعطل أو تحديد نوع العطل، وفقاً لطبيعة المهمة المطلوبة في نظام الصيانة التنبؤية.

يمثل هذا النموذج العصبي خطوة محورية في تطوير إطار صيانة تنبؤية متكامل، إذ يجمع بين قوة النماذج الشجرية في اكتشاف الأعطال الشائعة ومرونة الشبكات العصبونية في التعامل مع الأعطال النادرة، مما يعزز موثوقية النظام التنبؤي ويجعله مناسباً للتطبيق في البيئات الصناعية الحديثة.

3 النتائج

3.1 المرحلة الأولى: الكشف عن العطل

في المرحلة الأولى من النظام المقترح، تم التركيز على مهمة الكشف عن العطل في الآلة.

3.1.1 دراسة عطل الآلة باستخدام الميزات الخام

تم تدريب النماذج على متغير Machine Failure باعتباره المتغير الهدف وذلك بهدف التمييز بين حالتي "وجود عطل" و"عدم وجود عطل" في الآلة، تم تطبيق مجموعة من خوارزميات تعلّم الآلة على الميزات الخام لتقييم قدرتها على اكتشاف العطل بشكل أولي.

يوضح جدول 1 تقييم أداء هذه النماذج باستخدام مجموعة من المقاييس تشمل (Accuracy)، (Recall)، (Precision)، ومقياسF1 ، بالإضافة إلى تحليل مصفوفة الالتباس (Confusion Matrix).

وقد ساعد هذا التنوع في النماذج على فهم سلوك البيانات بشكل أعمق وتحديد النموذج الأنسب للانتقال إلى المرحلة الثانية، نتج مايلي:

جدول 1 تقييم أداء خوارزميات تعلم الآلة باستخدام الميزات الخام

Model	Accuracy	Recall	Precision	F1-Score
XGBoost Baseline	0.9875	0.6912	0.9216	0.7899
Gradient Boosting Baseline	0.9835	0.5882	0.8889	0.7080
Random Forest Baseline	0.9825	0.5441	0.9024	0.6789
Neural Network (MLP 64→32→16)	0.9785	0.5588	0.7451	0.6387
Decision Tree Baseline	0.9785	0.4706	0.8205	0.5981
SVM Baseline	0.9200	0.8529	0.2788	0.4203
Logistic Regression Baseline	0.8265	0.7794	0.1377	0.2340

على الرغم من أن جميع النماذج المحسوبة باستخدام الميزات الخام حققت مستويات مرتفعة من الدقة، إلا أنّ قيم كل من Recall و Precision كانت منخفضة نسبياً، مما يشير إلى ضعف قدرة النماذج على اكتشاف حالات العطل بشكل صحيح وتمييزها عن الحالات الطبيعية، وتبرز هذه النتائج الحاجة إلى إدخال ميزات مشتقة أكثر تعبيراً عن سلوك الآلة، بهدف تحسين قدرة النماذج على التقاط الأنماط المرتبطة بالعطل ورفع أداء مقاييس التصنيف الحساسة للفئات النادرة.

ولضمان بناء نماذج تنبؤية مستقرة وذات موثوقية عالية وقابلة للتطبيق في البيئات الصناعية تم التركيز على الأعطال الأربعة الأكثر تمثيلاً والأعلى ارتباطاً بالمتغيرات التشغيلية على الترتيب:

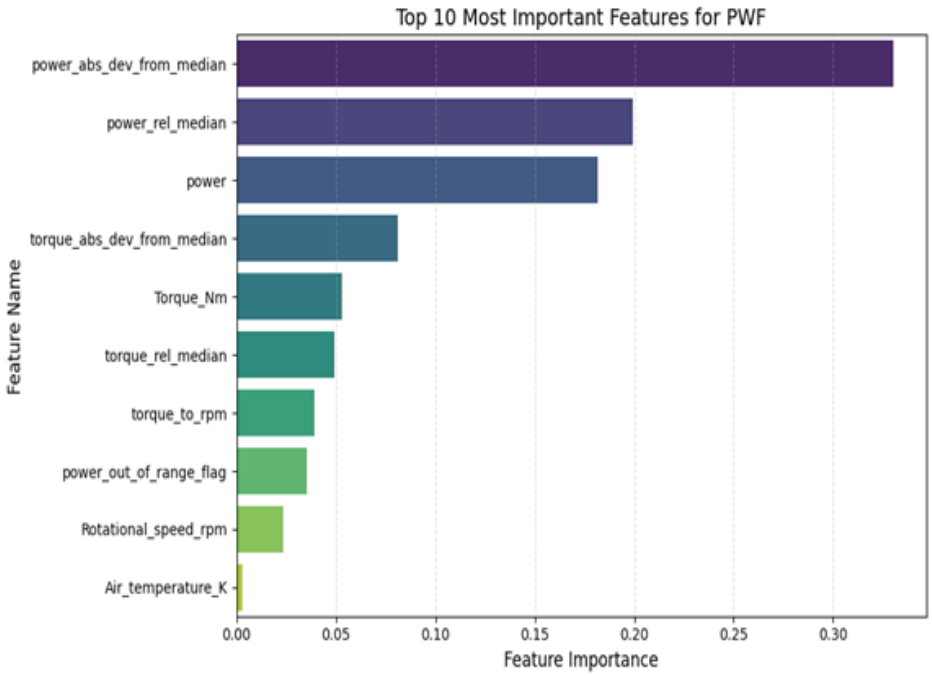
1. PWF
2. HDF
3. OSF
4. TWF

3.1.1.1 تحديد أهمية الميزات المؤثرة في كل عطل باستخدام Random Forest

تم الاعتماد على نموذج Random Forest نظراً لقدرته العالية على التقاط التفاعلات غير الخطية بين المتغيرات وتقدير أهمية كل ميزة ضمن سياق التصنيف، يتميز هذا النموذج بآلية داخلية لحساب أهمية الميزات، مما يتيح تحديد الميزات الأكثر ارتباطاً بكل نمط عطل، وبالتالي توجيه عملية هندسة الميزات نحو بناء تمثيل أكثر دقة لسلوك الآلة.

- **PWF**

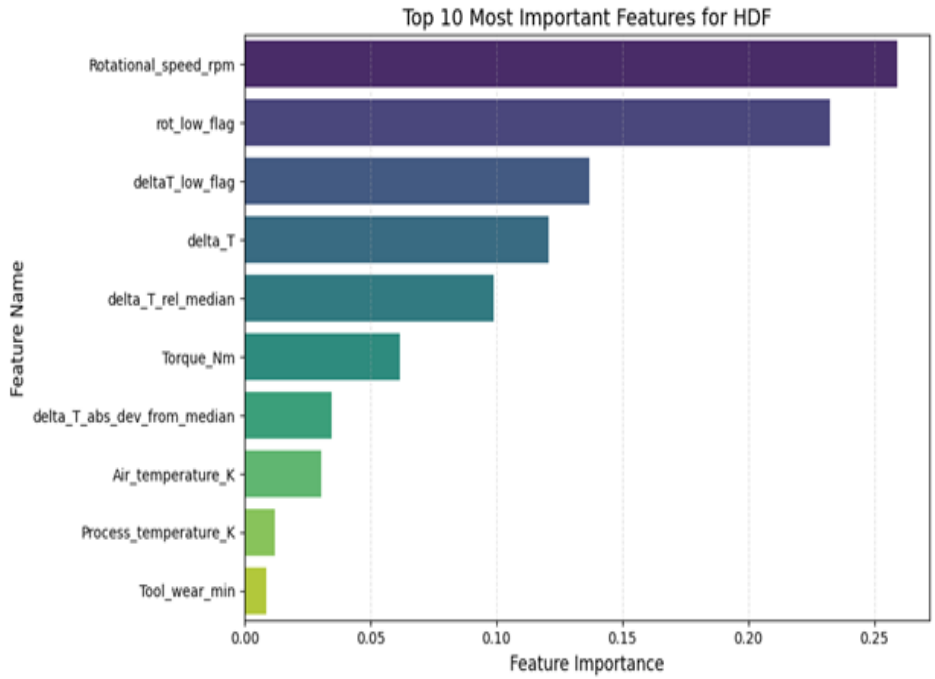
يبين شكل 2 ترتيب أهم الميزات لنمط العطل PWF نلاحظ أن الميزات المرتبطة بالقدرة Power مثل power_abs_dev_from_median و power_rel_median هي الأكثر مساهمة، ما يدل على أن انحرافات القدرة عن السلوك الطبيعي تلعب الدور الأكبر في كشف العطل،



شكل 2 ترتيب أهم الميزات لنمط العطل PWF

- **HDF**

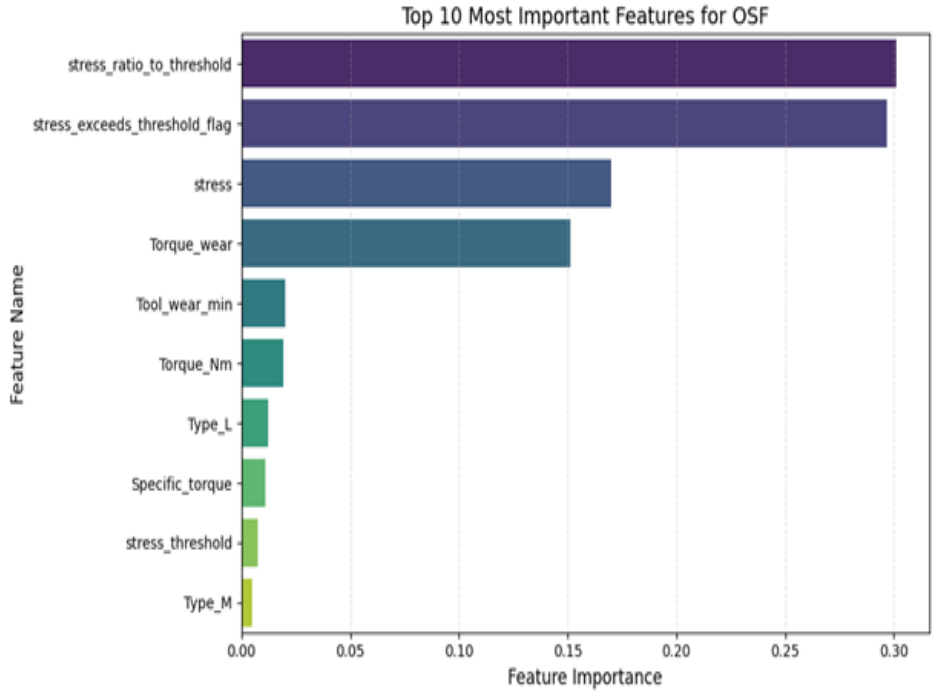
يوضـــح شـــكل 3 ترتيب أهم الميزات لنمط العطل HDF نلاحظ أن الميزات المرتبطة بانخفاض الســرعة الدورانية مثل Rotational_speed_rpm و rot_low_flag كانت الأكثر مساهمة، ما يدل على أن سلوك السرعة هو العامل الأبرز في كشف هذا النوع من الأعطال.



شكل 3 ترتيب أهم الميزات لنمط العطل HDF

- **OSF**

نلاحظ من الشكل 4 ترتيب أهم الميزات لنمط العطل OSF، أن الميزات المرتبطة بالإجهاد الميكانيكي مثل stress_ratio_to_threshold و stress_exceeds_threshold_flag هي الأكثر تأثيراً، ما يشير إلى أن تجاوز الإجهاد للحدود التصميمية هو العامل الأساسي في حدوث هذا العطل.



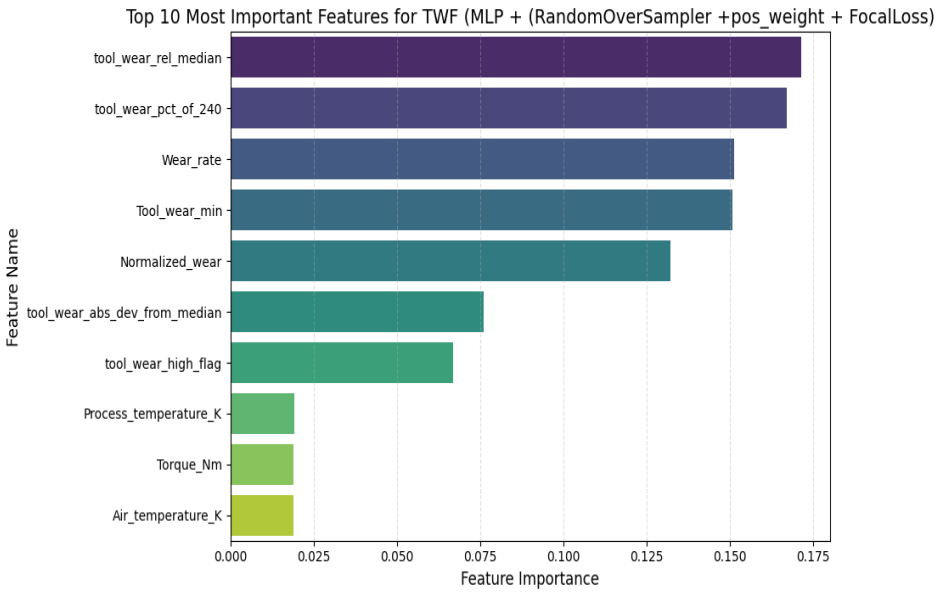
شكل 4 ترتيب أهم الميزات لنمط العطل OSF

- **TWF**

نظراً لندرة عطل TWF مقارنةً ببقية الأعطال، تم التعامل معه كنموذج فرعي مستقل ضمن بنية هجينة لضمان رفع حساسية الكشف عنه.

ولتحقيق ذلك، تم تدريب شبكة عصبونية متعددة الطبقات (MLP) مخصّصة لاكتشاف هذا العطل باستخدام مزيج من تقنيات معالجة اختلال التوزيع الطبقي.

يوضح شكل 5 أهم الميزات لنمط العطل TWF بعد معالجة عدم التوازن:



شكل 5 ترتيب أهم الميزات لنمط العطل TWF بعد معالجة عدم التوازن

3.1.2 دراسة عطل الآلة باستخدام الميزات المشتقة

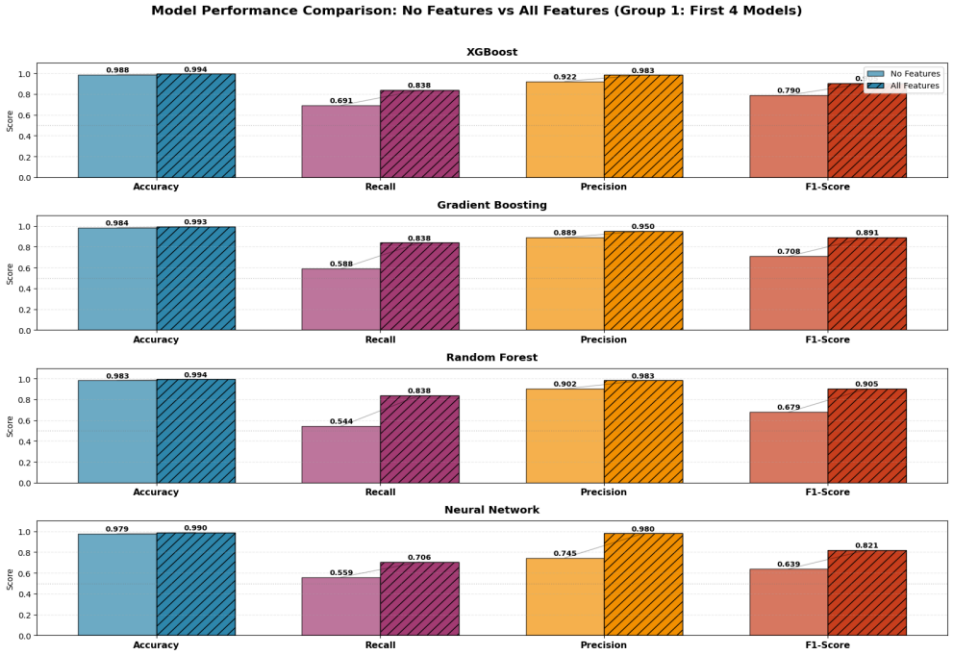
يبين جدول 2 نتائج تطبيق مجموعة من خوارزميات تعلّم الآلة على هذه الميزات المشـتقة بهدف تقييم قدرتها على اكتشـاف حالات العطل، وذلك من خلال تحليل أداء النماذج باسـتخدام مقاييس التقييم.

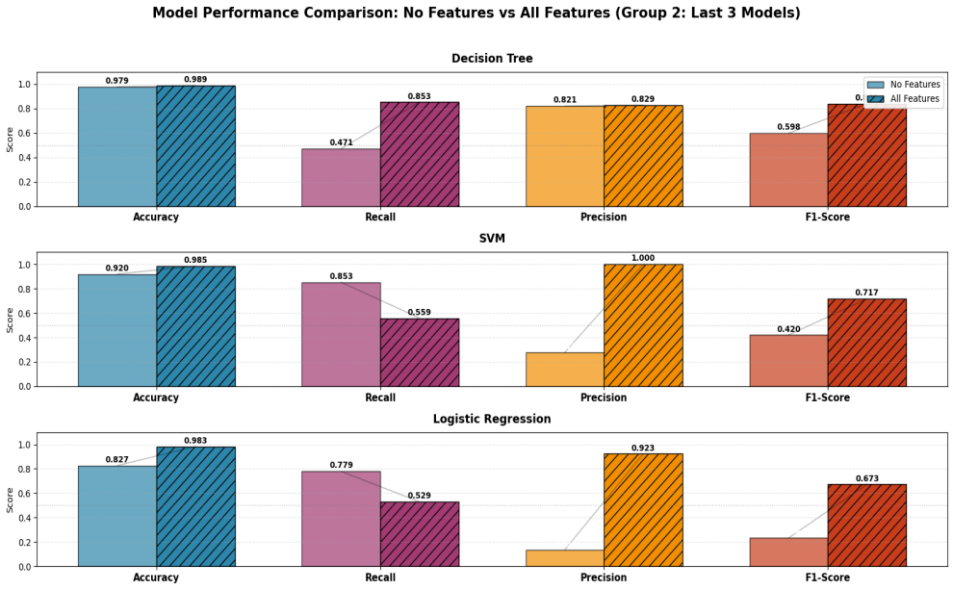
Model	Accuracy	Recall	Precision	F1-Score
XGBoost (Baseline)	0.9940	0.838235	0.982759	0.904762
Gradient Boosting (Baseline)	0.9930	0.838235	0.950000	0.890625
Random Forest (Baseline)	0.9940	0.838235	0.982759	0.904762
Neural Network (MLP 64→32→16)	0.9895	0.705882	0.979592	0.820513

Decision Tree (Baseline)	0.9890	0.852941	0.828571	0.840580
SVM (LinearSVC Baseline)	0.9850	0.558824	1.000000	0.716981
Logistic Regression (Baseline)	0.9825	0.529412	0.923077	0.672897

جدول 2 تقييم أداء خوارزميات تعلم الآلة باستخدام الميزات المشتقة

3.1.3 مقارنة أداء النماذج قبل وبعد استخدام الميزات المشتقة





شكل 6 مقارنة أداء النماذج قبل وبعد استخدام الميزات المشتقة

يوضح شكل 6 المقارنة بين أداء النماذج قبل وبعد استخدام الميزات المشتقة أثر هندسة الميزات في تحسين قدرة النماذج على اكتشاف حالات العطل، حيث أظهرت معظم النماذج تحسّناً ملحوظاً في مقاييس الاسترجاع وF1 Score بعد إدراج الخصائص المشتقة التي تعبّر بشكل أدق عن السلوك الميكانيكي للنظام.

ولكن على الرغم من أن استخدام الميزات المشتقة قد أدى إلى تحسين ملحوظ في دقة النماذج وارتفاع كبير في Precision وF1 Score، إلا أن قيمة Recall انخفضت في SVMوLogistic Regression ويُعزى هذا الانخفاض إلى أن الميزات المشتقة، رغم قدرتها على تعزيز الفصل بين الفئات، قد جعلت النموذج أكثر تحفظاً في تصنيف العينات على أنها حالات عطل، فالنماذج الخطية مثل SVM وLogistic Regression تعتمد على حدود قرار حادة، وعندما تصبح الميزات أكثر تمييزاً، فإن النموذج يميل إلى تضييق منطقة القرار الخاصة

بالفئة الإيجابية (العطل)، مما يقلل من عدد العينات المصنّفة كعطل ويؤدي بالتالي إلى انخفاض Recall.

3.2 المرحلة الثانية : تحديد نوع العطل

في هذه المرحلة، تم توحيد جميع أنواع الأعطال الجزئية في متغير واحد متعدد الفئات بهدف إنشاء هدف تصنيفي موحّد يسهل التعامل معه في نماذج تعلّم الآلة، ولتحقيق ذلك، جرى إنشاء عمود جديد باسم Machine failure كما هو موضح في جدول 3:

جدول 3 جدول ترميز فئات أعطال الآلة

Condition	Assigned Value
No Failure	0
TWF == 1	1
HDF == 1	2
PWF == 1	3
OSF == 1	4
RNF == 1	5

وبعد تنفيذ النماذج الأربعة الأولى على متغير Machine Failure المرمّز، كانت نتائج الأداء كما هي مبينة في جدول 4 :

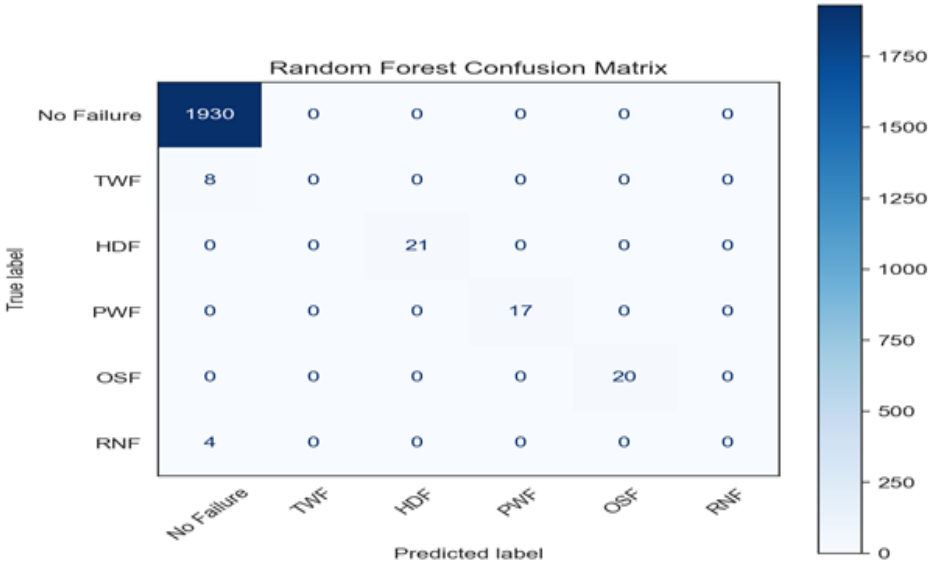
جدول 4 مقارنة أداء النماذج في مرحلة تحديد نوع العطل

Model	Accuracy	Recall	Precision	F1-Score
XGBoost Classifier	0.9940	0.9940	0.988037	0.991009
Gradient Boosting	0.9925	0.9925	0.987537	0.990005
Random Forest	0.9940	0.9940	0.988037	0.991009

Neural Network (MLP 64→32→16)	0.9880	0.9880	0.981427	0.983787

تُظهر نتائج النماذج أن أداء خوارزميات تعلّم الآلة ظلّ مرتفعاً عند التعامل مع متغير Machine failure متعدد الفئات وتؤكد هذه النتائج أن عملية الترميز لم تُضعف أداء النماذج، بل مكّنتها من التعرّف على أنماط الأعطال المختلفة بكفاءة عالية، مما يجعلها مناسبة للتنبؤ بنوع العطل بشكل أكثر دقة وموثوقية.

يعرض شكل 7 مصفوفة الالتباس الخاصة بنموذج Random Forest بعد تطبيقه على مهمة تصنيف نوع العطل باستخدام متغير Machine Failure المرمّز.



شكل 7 مصفوفة الالتباس لنموذج الغابات العشوائية (تصنيف الأعطال)

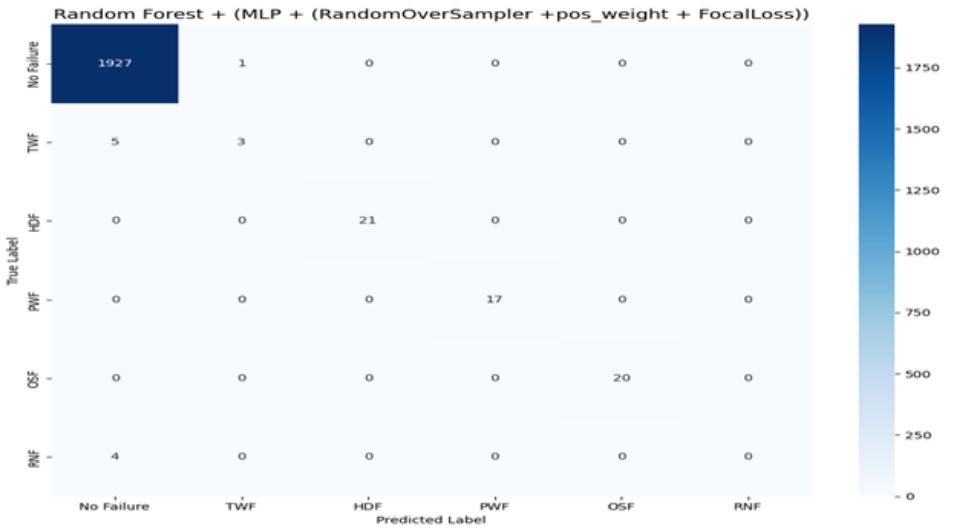
توضح المصفوفة أن النموذج نجح بشكل ممتاز في تصنيف الحالات السليمة، حيث تعرّف على جميعها دون أخطاء، كما أظهر دقة عالية في تصـنيف أعطال HDF و PWF وOSF، إذ تم التعرّف عليها بالكامل.

في المقابل، لم يتمكّن النموذج من التعرّف على أي حالة من حالات TWF ، حيث تم تصنيفها جميعاً ضمن الفئة السليمة، مما يشير إلى ضعف النموذج في التعامل مع الأعطال النادرة.

ولمعالجة مشكلة ضعف النماذج في التعرّف على عطل TWF تم استخدام نهج هجين كما في المرحلة الاولى وقد أسـهم هذا النهج الهجين في رفع حسـاسـية النظام تجاه عطلTWF دون التأثير على أداء النماذج في الفئات الأخرى، مما أدى إلى بنية أكثر توازناً وموثوقية في تحديد نوع العطل [17].

أما عطل RNF، فقد تم اسـتبعاده من النموذج المخصّـص للأعطال النادرة نظراً لغياب نمط تشـغيلي واضـح يمكن للنموذج تعلّمه، إضـافةً إلى ندرة العينات المرتبطة به، مما يجعل بناء نموذج مسـتقل له غير مسـتقر وغير قابل للتعميم، ويؤكد هذا الواقع محدودية القدرة على التنبؤ بهذا النوع من الأعطال، نظراً لعدم توفر خصـائص مميزة أو سـلوك ميكانيكي ثابت يمكن استغلاله نمذجياً.

يعرض شـكل 8 نتائج النموذج الهجين الذي يجمع بين Random Forest و MLP ، مع استخدام تقنيات معالجة عدم التوازن لتحسين التعرّف على الأعطال النادرة



شكل 8 مصفوفة الالتباس النهائية لنموذج التصنيف الهجين

تُظهر المصفوفة تحسناً ملحوظاً في اكتشاف عطل TWF مقارنة بالنموذج السابق، حيث تم تصنيف ثلاث حالات بشكل صحيح بعد أن كانت جميعها تُخطئ سابقاً، كما حافظ النموذج على دقة عالية في تصنيف الأعطال الشائعة HDF وPWF وOSF، دون أي تراجع في الأداء، وتُبرز هذه النتائج فعالية النهج الهجين في تعزيز حساسية النموذج تجاه الأعطال النادرة دون التأثير على دقته في الفئات الأخرى.

4 مقارنة نتائج النموذج المقترح مع الدراسات السابقة

بعد دراسة وتحليل مقاييس الأداء الخاصة بالنموذج المقترح في هذا البحث، تم إجراء مقارنة منهجية بين نتائج النموذج الحالي ونتائج مجموعة من الدراسات المرجعية الحديثة [5]،[6]،[7]،[1] في مجال الصيانة التنبؤية كما هو موضح في جدول 5.

جدول 5 مقارنة مع الأعمال ذات الصلة

الدراسة	نوع البيانات	الميزات المستخدمة	معالجة عدم التوازن	أفضل نموذج	تصنيف الأعطال (Multi-Class)
5	بيانات حساسات حقيقية من آلات غزل صناعية	ميزات خام	غير مستخدمة	Random Forest	✗
6	بيانات تشغيل كهربائية حقيقية لمحركات كهربائية	ميزات خام	Oversampling بسيط	XGBoost	✗
7	AI4I 2020	ميزات إحصائية بسيطة	غير مستخدمة	SVM	✗
1	AI4I 2020	ميزات خام	Class Weights	Gradient Boosting	✓
Our Work	AI4I 2020	ميزات خام و ميزات مشتقة	Oversampling + pos_weight + Focal Loss	MLP - XGBoost - Gradient Boosting -	✓

				Random Forest	

تُبرز المقارنة بين النموذج المقترح والدراســـات المرجعية الأربع تفوق المنهجية المعتمدة في هذا البحث، والتي تجمع بين هندســـة الميزات المتقدمة، ومعالجة عدم التوازن، واســـتخدام نماذج هجينة، فقد أظهرت الدراســـات الســـابقة اعتماداً كبيراً على الميزات الخام والنماذج التقليدية دون معالجة فعّالة لعدم التوازن، مما أدى إلى انخفاض واضـح في مؤشـرات Recall و F1-Score للفئات النادرة رغم ارتفاع الدقة العامة.

كما تفوقت الشبكة العصبونية المخصّصة على جميع النماذج المستخدمة في الدراسات السابقة من حيث قدرة اكتشـــاف الأعطال النادرة، مما يؤكد فعالية التصـــميم الهجين المقترح وملاءمته للتطبيقات الصناعية التي تتطلب دقة عالية في رصد حالات العطل.

5 خاتمة

تم في هذه الدراسـة تطوير نموذج صـيانة تنبؤية قادر على التعرّف على الأعطال التشـغيلية في آلات التفريز بدقة عالية، مع قدرة مميزة على اكتشـاف الأعطال النادرة مثلTWF . وقد سـاهم دمج هندسـة الميزات مع تقنيات معالجة عدم التوازن في تحسـين مؤشـرات الأداء بشـكل واضـح، بينما أظهر النموذج الهجين الذي يجمع بين النماذج الشـــجرية والشـــبكة العصـــبونية تفوقاً في تصنيف الأعطال الشائعة والنادرة على حد سواء، وتؤكد النتائج أن المنهجية المقترحة توفر إطاراً عملياً وفعالاً للصـــيانة التنبؤية، يدعم تقليل التوقفات غير المخططة ويعزز موثوقية المعدات في البيئات الصناعية.

المراجع

1. Zhao, W., Ding, J., Huang, X., & Zhang, Y. (2025). Research on milling machine predictive maintenance based on machine learning and SHAP analysis in intelligent manufacturing environment (arXiv:2512.01205v1). arXiv.

2. Kumar, R., et al. (2023). Impact of machine downtime on modern manufacturing systems. Procedia CIRP.

3. Carvalho, T., et al. (2023). Predictive maintenance in Industry 4.0: Recent developments and challenges. Sensors.

4. Zhao, W., et al. (2025). Machine learning based fault diagnosis in CNC milling machines. Applied Sciences.

5. Akyaz, T., & Engin, D. (2024). Machine learning-based predictive maintenance system for artificial yarn machines. IEEE Access, 12, 125446–125461.

6. Baradaran, A. H. (2025). Predictive maintenance of electric motors using supervised learning models: A comparative analysis (arXiv:2504.03670). arXiv.

7. Waghulde, R. R., Rai, R. K., Chadhar, R. M., Rane, M., & Yadav, V. (2025). Evaluating machine learning and deep learning models for predictive maintenance: A study using the AI4I 2020 dataset. Journal Name.

8. AI4I 2020 Predictive Maintenance Dataset- UCI Machine Learning Repository (2020)

9. Wang, Z., Liu, Q., & Zhao, R. (2022). Data imbalance challenges in predictive maintenance: A comprehensive review. IEEE Transactions on Industrial Informatics, 18(8), 5234–5248.

10. https://www.sanfoundry.com/data-preprocessing-ml-cleaning-normalization-encoding

11. Tremblay, R. (2025). Understanding equipment failure behavior: Types, patterns, and maintenance strategies. Spartakus.

12. Z. M. Çınar et al., "Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0," Sustainability, vol. 12, art. 8211, Oct. 2020.

13. Buabeng, A. Simons, N. K. Frempong, and Y. Y. Ziggah, "Hybrid Intelligent Predictive Maintenance Model for Multiclass Fault Classification," Soft Computing, vol. 28, pp. 8749–8770, Jul. 28, 2023.

14. Chugani, V. (2025). A concise guide to overfitting. Statology.

15. Crudu, A., & MoldStud Research Team. (2025). Mitigating vanishing gradients with advanced activation functions — A comprehensive guide.

16. Gu, J., Li, C., Liang, Y., Shi, Z., & Song, Z. (2024). Exploring the frontiers of Softmax: Provable optimization, applications in diffusion model, and beyond (arXiv preprint arXiv:2405.00000). arXiv.

17. J. Wiercioch, "Hybrid predictive maintenance model – study and implementation example," Production Engineering Archives, vol. 30, no. 3, 2024.